



Comparison of habitat models for scarcely detected species



Auriane Virgili^{a,*}, Mélanie Racine^b, Matthieu Authier^b, Pascal Monestiez^{a,c}, Vincent Ridoux^{a,b}

^a Centre d'Etudes Biologiques de Chizé—La Rochelle, UMR 7372 CNRS—Université de La Rochelle, Institut du Littoral et de l'Environnement, 17000 La Rochelle, France

^b Observatoire PELAGIS, UMS 3462 CNRS—Université de La Rochelle, Systèmes d'Observation pour la Conservation des Mammifères et des Oiseaux Marins, 17000 La Rochelle, France

^c BioSP, INRA, 84914 Avignon, France

ARTICLE INFO

Article history:

Received 11 July 2016

Received in revised form

19 December 2016

Accepted 21 December 2016

Available online 10 January 2017

Keywords:

Habitat modelling

Scarce species

GAM

GLM

MaxEnt

Poisson

Negative binomial

Tweedie

Zero-inflated Poisson

ABSTRACT

When performing habitat models, modellers have to choose between presence-absence and presence-only models to estimate the habitat preferences of a species. Primarily, this choice depends on the data that are available and whether effort data are recorded in parallel to sighting data. For species that are rare or scarce, the models have to address a great number of zeros (*i.e.*, no animal seen) that weakens the ability to make sound ecological inferences. We tested two types of habitat models (presence-absence vs. presence-only) to determine which type best dealt with datasets containing an excess of zeros, and we applied our models to a sighting dataset that included the common (*Delphinus delphis*) and striped (*Stenella coeruleoalba*) dolphin (approximately 92% zeros). We used two types of presence-absence models (Generalised Additive models – GAMs, Generalised Linear Model – GLM) and one presence-only model, a MaxEnt model, and we used various criteria to compare these models (*i.e.*, AIC, deviances, rootograms and distribution patterns predicted by the models). Overall, we observed that the presence-absence models made better predictions than the presence-only model. Among the presence-absence models, the GAM with a Negative Binomial distribution was better at predicting small delphinids habitats, even though the GAM with a Tweedie distribution exhibited similar results. However, the zero-inflated Poisson distributions exhibited less convincing results and was contrary to what was expected. Finally, despite 92% zeros, our dataset was not zero-inflated. Our study demonstrates the importance of selecting appropriate models to make reliable predictions of habitat use for species that are rare or scarce.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Identifying habitat needed and used by species is important for wildlife management and conservation (Cañadas *et al.*, 2005; Bailey and Thompson, 2009). One means of identifying habitat is with statistical models that correlate the spatial distributions of animal sightings with environmental inputs (Austin, 2002; Guisan and Thuiller, 2005; Redfern *et al.*, 2006). Such models allow the habitat of a species and presence to be estimated. They also allow for predictions in areas that have not been previously surveyed (Segurado *et al.*, 2004).

Species distribution models have recently undergone rapid development and have been used for diverse applications (*e.g.*, Elith

et al., 2006; Elith and Leathwick, 2009; Mannoce *et al.*, 2014a, 2014b, 2015). There are generally two categories of habitat models: presence-absence and presence-only models; the chosen model depends on the type of data used and, notably, whether effort data are recorded in parallel to sighting data (Guisan and Zimmermann, 2000).

The first group of models requires presence and effort data that are recorded during planned surveys, where each on-effort sighting represents a detection of the target species. Such presence-absence models include, among others, generalised linear models (GLM), generalised additive models (GAM), regression trees analyses such as boosted regression trees (BRT) (Guisan and Zimmermann, 2000; Brotons *et al.*, 2004), or occupancy models (MacKenzie *et al.*, 2002). Some of these models allow estimating detection probability, and consequently, prediction of habitat suitability of a species (Gormley *et al.*, 2011). They also allow functional relationships to be fitted between species locations and local environmental conditions (Guisan and Zimmermann, 2000). The models of the second group

* Corresponding author. Postal address: 2 rue Olympe de Gouges, 17000 La Rochelle, France.

E-mail address: auriane.virgili@univ-lr.fr (A. Virgili).

only require detection data, such as opportunistic data, where the absence data are missing because effort data were not documented and non-detection data are not prospected and informed (Hirzel et al., 2002). These include Ecological Niche Factor Analysis (ENFA) or Maximum Entropy Modelling (MaxEnt), and allows for the identification of potentially suitable sites by evidencing the environmental conditions that are similar to the sites where animals were recorded (Elith et al., 2006). Nevertheless, the accuracy of presence-only model outputs is conditional on random or representative sampling of the habitat at the data collection stage (Yackulic et al., 2013). Presence-only data are the default option when data on absence (that is effort data) are not available (Zaniewski et al., 2002).

Except for presence-only models, which do not consider the zeros, choosing among presence-absence models might be difficult depending on the studied species, particularly when focusing on scarcely detected species, because of the inherent difficulty of models to accommodate a large number of absences. Due to restricted habitat range, low density and poor detection even in favourable habitats (Martin et al., 2005), the number of absences in some datasets (*i.e.*, the zeros) can be large. True (or structural – the taxon is really absent from an area), and false (or sampling – the taxon is present but poorly detected) absences become particularly challenging to tell apart (Ridout et al., 1998).

Due to their discrete probability distribution, count data are basically modelled with a Poisson regression, but when compared to this Poisson distribution, ecological data are often over-dispersed (*i.e.*, the variance is greater than the mean) and require specific treatment to avoid biased results (Ridout et al., 1998; Dobbie and Welsh, 2001). Failure to accommodate over-dispersion leads to the selection of a model that is more complex than necessary (Richards, 2008), where the model does not generalise outside the sample used to calibrate it. One reason for over-dispersion that has attracted much attention is zero-inflation (Deng and Paul, 2005), where a large abundance of zeros in a dataset needs to be adequately analysed to prevent model misspecification and misleading ecological conclusions due to the under- or over-estimation of some functional relationships. Too many zeros can also increase biases and uncertainties in the estimated model parameters (MacKenzie et al., 2002; Martin et al., 2005). Hence, habitat modellers face two main issues: first, they have to define if their data are under-, equi- or over-dispersed, and second, depending on their data, they have to find an appropriate model for the dispersion (for example, zero-inflated models).

The selection of a good enough (that is accurate) model is critical for habitat models to fulfil their potential for management and conservation purposes. Habitat models can reveal areas of high densities of organisms; they can help to define or confirm key areas of conservation in order to meet stakeholder expectations (Cañadas et al., 2005). This is even more important when focusing on scarcely detected species because these areas of high densities are more difficult to identify.

Consequently, the aim of our study was to help habitat modellers find an appropriate model when working with data with many zeros. To do that, we compared the predictive performance of both presence-absence and presence-only models and tested their ability to address an apparently zero-inflated dataset. We used a small delphinids sightings dataset; which pooled the common *Delphinus delphis* and the striped *Stenella coeruleoalba* dolphin. These data include approximately 92% zeros. Small delphinids show distribution patterns that are easily identified by habitat modelling, and thus they allow a comparison of different models. They are typical top predators in that they are sparsely distributed *in natura*. Associated datasets are characterised by the presence of many zeros even within favourable habitats (Redfern et al., 2006). However, they provide sufficient data to fit various distribution models and

statistically compare their outputs. Using this dataset, we tested different models: GAMs with a Poisson, a Negative Binomial, a Tweedie and a zero-inflated Poisson distribution; a GLM with a zero-inflated Poisson distribution and a presence-only model; the MaxEnt model. Due to their ability to model separately the absences and the presences (Lambert, 1992), we assumed *a priori* that a zero-inflated Poisson model would perform best. However, the Negative Binomial and Tweedie distributions can also provide good fits (Warton, 2005; Dunn and Smyth, 2005; Lindén and Mantyniemi, 2011). In addition, with its multiple applications (Yackulic et al., 2013), including those by managers, and its ability to take into account the complex interactions between response and predictor variables (Elith et al., 2006, 2011; Phillips et al., 2004; Phillips and Dudík, 2008), the MaxEnt model appears to be a relevant tool for modelling habitats of rare species (Wisiz et al., 2008). Therefore, we also tested the model to assess its efficiency. This study aims to pragmatically answer some questions commonly asked by habitat modellers, such as those regarding the effective zero-inflation of their data and the relevance of the chosen model depending on their data.

2. Materials and methods

2.1. Datasets

2.1.1. Aerial surveys and data collection

The small delphinids sighting data were recorded during the SAMM survey (*Suivi Aérien de la Mégafaune Marine*; Aerial Census of Marine Megafauna), which was dedicated to the observation of marine megafauna and conducted in the northeast Atlantic Ocean and the northwest Mediterranean Sea (Laran et al., *in review*; Lambert et al., *in press*). In the present work we focused on data collected in the summer of 2012 in the entire English Channel and the Bay of Biscay from the tip of Brittany to the Dover Strait in the north, and to the Spanish coast in the south (Fig. 1). The survey was carried out from mid-May to early August along 31,427 km of transect lines. A standard methodology for cetacean surveys was applied (Hammond et al., 2013) using twin-engine high-wing aircrafts equipped with bubble windows. The flights followed a zig-zag pattern, at a speed of 167 km/h and an altitude of 183 m. Observation conditions (Beaufort sea state, turbidity, cloud cover and glare severity) and sightings with group size were recorded following a line-transect methodology (Buckland et al., 2001). This implies that the angle between every sighting and the track line was recorded to estimate the Effective Strip Width (ESW; see the small delphinids detection function and estimated ESW in Laran et al., *in review*).

The common and striped dolphins were pooled because it was most often impossible to tell apart the two species from the plane. During the survey, 277 sightings of small delphinids were recorded in good observation conditions, corresponding to 14,477 individuals (Fig. 1).

2.1.2. Environmental predictors

To model the relationships between small delphinids and their environment, we used eight environmental predictors (Lambert et al., *in press*; Virgili et al., *in review*), of which there were two physiographic variables (depth and slope) and six oceanographic variables (mean, variance and gradient of Sea Surface Temperature—SST, mean and standard deviation of Sea Surface Height—SSH, and the maximum velocity of tidal currents (Table 1)). All oceanographic variables were computed at a seven day resolution, *i.e.*, averaged over 6 days prior to the sampled day. Physiographic variables are static and relate to the bathymetry, and oceanographic variables are dynamic and describe water masses.

Download English Version:

<https://daneshyari.com/en/article/5742348>

Download Persian Version:

<https://daneshyari.com/article/5742348>

[Daneshyari.com](https://daneshyari.com)